IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA	)	
	)	
v.	)	Criminal No. 19-369
	)	
LAFON ELLIS	)	

Declaration of Nathaniel Adams

I, Nathaniel Adams, declare that I have personal knowledge of the following, and if called upon to do so, could and would testify competently to the matters contained herein.

I. Qualifications

1. I have a Bachelor of Science degree with a major in Computer Science from Wright State University (Dayton, Ohio). I am enrolled in the Graduate School at Wright State University, pursuing a Master of Science degree in Computer Science.

2. I am employed in Fairborn, Ohio as a Systems Engineer at Forensic Bioinformatic Services, Inc., where I have worked for seven years. My duties include analyzing electronic data generated during the course of forensic DNA testing; reviewing case materials; reviewing laboratory protocols; and performing calculations of statistical weights, including custom simulations for casework and research projects. I actively use, develop, and maintain a number of software programs to assist with these efforts.

3. In 2014 I attended a week-long workshop on interpreting forensic DNA mixtures using emerging software solutions, including TrueAllele®.

4. As part of my work, I am regularly involved in reviews of probabilistic genotyping (PG) analyses, including TrueAllele®, in criminal cases. I regularly present at academic and legal education conferences and workshops on forensic DNA topics, including probabilistic genotyping. I have testified as an expert about probabilistic genotyping software in state and federal courts in

Exhibit 1

the United States and in state courts in Australia and Canada.

5. I have reviewed non-public software development materials, including source code, for probabilistic genotyping systems STRmix™ and FST used in criminal cases in state and federal courts in the United States and Australia. Due to non-disclosure agreements and protective order acknowledgements that I signed, I am not allowed to discuss the findings of several of these reviews outside of their cases. Thus, the scope of this declaration is limited to information that is publicly available or has been provided in this case.

II. Overview

6. I have been asked by attorney Khasha Attaran to address concerns regarding the development and use of the TrueAllele® probabilistic genotyping software system in the case of *United States v. Lafon Ellis.* Materials I have received in this case include a DVD of TrueAllele® materials, a subpoena for additional materials filed on April 29, 2020, an unsigned declaration by Mark Perlin dated April 2020, and the government's motion to quash subpoena filed June 16, 2020.

7. These concerns include the appropriate levels of confidence to assign the likelihood ratios ("match scores") calculated by TrueAllele® and reported in this case based on software engineering principles, standards, and practices.

8. There is very little information publicly available about the development of the TrueAllele® software, including materials requested in the subpoena such as software engineering documentation or the source code.

9. I am not aware of any public claims by the owners of TrueAllele® (Cybergenetics) that it has been constructed, verified, or validated in accordance with any software engineering standard document, either internal or external to Cybergenetics.

10. Three forensic DNA organizations have issued guidance for the validation of probabilistic genotyping software: the United States Scientific Working Group on DNA Analysis Methods (SWGDAM) in 2015[1]; the International Society for Forensic Genetics (ISFG) in 2016[2]; and the United Kingdom Forensic Science Regulator (FSR) in 2018[3].

11. SWGDAM 2015 describes the interdisciplinary nature of probabilistic genotyping due to its use of "biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios," but does not describe software engineering practices to use during the development or validation of probabilistic genotyping software.

12. ISFG states, "International industry standards apply to software validation, verification and test documentation. These standards can be simplified and extrapolated to forensic genetics." ISFG's guidance cites the Institute of Electrical and Electronics Engineers (IEEE) Standard 1012-2012, *Standard for System and Software Verification and Validation*[4] as an example of a relevant existing software standard.

13. FSR states that specific international standards commonly used in forensic science "provide little by way of guidance regarding software validation." FSR's guidance document cites and discusses additional standards general to software development efforts such as Standard 12207, *Systems and software engineering – Software life cycle processes*. Standard 12207 was

---

[1] Sci. Working Grp. on DNA Analysis Methods, *Guidelines for the Validation of Probabilistic Genotyping Systems* (2015).

[2] M. D. Coble et al., *DNA Commission of the International Society for Forensic Genetics: Recommendations on the Validation of Software Programs Performing Biostatistical Calculations for Forensic Genetics Applications*, 25 Forensic Sci. Int'l: Genetics 191 (2016).

[3] U.K. Forensic Sci. Regulator, *Software Validation for DNA Mixture Interpretation: FSR-G-223*, no. 1 (2018).

[4] Inst. of Elec. & Electronics Eng'rs, *IEEE Std. 1012-2012: IEEE Standard for System and Software Verification and Validation* (2012).

jointly adopted by the International Organization for Standardization (ISO), the International Electronic Commission (IEC), and IEEE.

14. ISO is an international organization of national standards bodies. IEC is a similar organization specializing in electronics and software. IEEE is a professional electrical, computer, systems, and software engineering organization with over 400,000 members.[5] IEEE's Standards Association regularly publishes standards for the development and use of computer software, equipment, and systems. Many IEEE standards have been adopted by ISO and IEC. IEEE standards have also been adopted by US governmental agencies including the Food and Drug Administration,[6] the Department of Defense,[7] and the Nuclear Regulatory Commission.[8]

15. The United States National Institute of Standards and Technology (NIST) has begun recognizing standards specific to forensic science.[9] Presently there is no NIST-recognized standard specific to the development or validation of probabilistic genotyping software.

16. Without domain-specific software standards, general software engineering industry standards such as IEEE Std 1012-1012 can define and describe good practices for the development and validation of complex software.

---

[5] Inst. of Elec. & Electronics Eng'rs, *IEEE – About IEEE*, https://www.ieee.org/about/index.html (last accessed July 9, 2020).

[6] Food & Drug Admin., *General Principles of Software Validation; Final Guidance for Industry and FDA Staff* (2002).

[7] U.S. Dep't of Def., Modeling & Simulation Coordination Office, *Modeling & Simulation Verification, Validation, & Accreditation Recommended Practices Guide Core Document* (2011).

[8] U.S. Nuclear Reg. Comm'n, *Regulatory Guide 1.168 - Verification, Validation, Reviews, and Audits for Digital Computer Software Used in Safety Systems of Nuclear Power Plants* (1997).

[9] U.S. Nat'l Inst. of Standards & Tech., *OSAC Registry Approved Standards*, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/osac-registry-approved-standards (last accessed July 9, 2020).

III. Confidence in DNA mixture interpretation software systems

17. Likelihood ratios are inherently based in uncertainty. Without "ground truths" against which a calculation's result can be compared, other methods must be used to establish confidence in the accuracy of statistical software operations.

18. I disagree with the government's statement in the motion to quash subpoena that "[t]here is no genuine controversy as to the validity and reliability of the TrueAllele method." TrueAllele® is a complex software system whose output required 40 days' worth of computer time to compute in this case.[10] Confidence in these outputs should derive from confidence in the operations executed by TrueAllele®. Confidence in the operation of any software is determined through assessments of its quality as a software system.

19. IEEE proffers a definition of software quality as the "degree to which a product or process meets established requirements; however, quality depends upon the degree to which those established requirements accurately represent stakeholder needs, wants, and expectations."[11,12]

20. Assessments of degrees to which software conforms to requirements, needs, wants, and expectations are generally made under the banner of verification and validation (V&V).

21. ISFG helpfully describes validation as "is it doing the right thing?" and verification as "is it doing the thing right?"[13]

---

[10] Cybergenetics, *Case Packet* – Commonwealth of Pennsylvania v. Lafon Ellis, 1-124 (Dec. 5, 2019).

[11] P. Bourque & R. E. Fairley, Eds., *Guide to the Software Engineering Body of Knowledge v3.0* (2014).

[12] Inst. of Elec. & Electronics Eng'rs, *IEEE Std 730-2014: IEEE Standard for Software Quality Assurance Processes* (2014).

[13] IEEE Std 1012-2012 describes verification and validation in greater depth, including their overlap on evaluating how well a piece of software conforms to its requirements (i.e., intended behaviors): "The Verification Process provides objective evidence for whether the products perform the following: a) Conform to requirements (e.g., for correctness, completeness,

22. IEEE Std 1012-2012 lists software V&V processes, culminating in the 26-page "Table 1c—V&V tasks, inputs, and outputs." For TrueAllele®, these V&V tasks would include the creation and evaluation of software development materials requested in paragraph 5 of the amended subpoena. Source code, for example, is listed in Table 1c as a required input for more than ten V&V tasks. TrueAllele® software requirements specifications, software design documents, and software test plans and reports are not public and would be required for the V&V tasks described in IEEE Std 1012-2012.

23. Some IEEE Std 1012 V&V material 'inputs' requested in paragraph 5 of the amended subpoena have not been mentioned in published articles about TrueAllele® or in the validation study summaries included on the TrueAllele® DVD provided in this case. These articles and non-peer-reviewed summaries do not claim conformance with processes described in software engineering standards documents. Consequently, they should not be considered evidence of completing software engineering V&V.

24. It is appropriate to question which types of materials used for V&V processes do exist for TrueAllele® and how these materials have been or could be used for TrueAllele® V&V efforts. V&V tasks described by IEEE Std 1012-2012 cannot be completed if they do not involve reviews and evaluations of the required input materials, regardless of any peer-review process used.

25. IEEE Std 1012-2012 describes a spectrum of V&V task granularity that should be

---

consistency, and accuracy) for all activities during each life cycle process; b) Satisfy the standards, practices, and conventions during life cycle processes; c) Successfully complete each life cycle activity and satisfy all the criteria for initiating succeeding life cycle activities (i.e., builds the product correctly). The Validation Process provides evidence for whether the products perform the following: Satisfy system requirements allocated to the products at the end of each life cycle activity; Solve the right problem (e.g., correctly model physical laws, implement business rules, and use the proper system assumptions); Satisfy intended use and user needs in the operational environment (i.e., builds the correct product)."

conducted based on a software system's integrity level. The proposed levels range from 1 (lowest) to 4 (highest). Levels 3 and 4 are assigned to systems where errors in the software's operation can lead to "catastrophic consequences," defined as "Loss of human life, complete mission failure, loss of system security and safety, or extensive financial or social loss."

26. IEEE Std 1012-2012 describes degrees of managerially, technically, and financially independent verification and validation (IV&V) efforts as "appropriate" or "generally required" for software systems of integrity levels 3 or 4, respectively.

27. Most TrueAllele® peer-reviewed articles and many non-peer-reviewed validation summaries include Cybergenetics personnel as co-authors. If TrueAllele® should be expected to adhere to V&V tasks for integrity level 3 or 4, it is also appropriate to question the degree to which TrueAllele® V&V efforts involved independent personnel, especially V&V efforts involving non-public materials.

28. I also disagree with the government's statement, "In addition, TrueAllele's reliability was established on the evidence in this case. The report and its supporting case packet provided to the defense described the system's sensitivity, specificity and reproducibility on the DNA evidence." Qualitatively "reproducible" results can signify precision (how close results are to each other), but not accuracy (how close a result is to the true value). Sensitivity rates are the ability of a system to detect true positives (known contributors to a sample), which are inherently unknown for casework samples.

29. The government proceeds to state, "The case packet gives the data and parameter inputs used in running the program in the case. The packet also includes a case-specific minivalidation study of reported TrueAllele match statistics, measuring match specificity by comparison with non-contributor genotypes. Source code is not needed to understand or interpret these materials."

Source code is a required input of multiple V&V tasks and consequently would be used in a review of TrueAllele® V&V efforts, which should inform us as to the confidence we should place in the outputs of TrueAllele®, including "match statistics" and specificity rates or distributions. The outputs themselves do not possess an inherent ability to prove their own correctness.

30. The government also states, "Additional discovery material for this case was provided on an optical disc. The DVD contains documents related to TrueAllele's reliability, such as background reading, over thirty validation studies and publications, regulatory approvals, general acceptance, and admissibility rulings." As stated above, these documents do not demonstrate conformance to standardized V&V processes described in software engineering standards documents.

31. The government states, "There are tutorial videos that describe TrueAllele methods and explain how the system works, as well as continuing legal education talks. The VUIer™ software for reviewing TrueAllele results is provided (with both Windows and Macintosh installers), along with instructions and user manuals. Case-specific files (data, reports, PowerPoint, case packet, VUIer input) are disclosed, enabling a thorough expert review. Source code is not needed to access these materials, read the files, use the executable VUIer software, or examine the computer results." V&V efforts evaluate conformance of the operations undertaken by TrueAllele® during its hours- or days-long calculations with the methods described in its articles, summaries, tutorials, and case-specific materials. Formalized, enumerated software requirement specifications for TrueAllele®, including methods for qualification testing the system against the requirements, are expected to be the most discrete, objective, and falsifiable descriptions of intended system operations. Through testing, comparison of the software requirement specifications to the human-readable computer instructions (source code), claims of conformance of the actual system to its

intended and advertised behaviors can be evaluated.

IV. The need for Source code and intended review

32. In 2016 my company and Dr. Eli Shapiro were hired in *United States v. Kevin Johnson* (S. Dist. of NY 1:15-cr-00565) to review the New York City Office of the Medical Examiner's (OCME) probabilistic genotyping software system Forensic Statistical Tool (FST). Like TrueAllele®, FST operates in a client-server architecture including database storage of analysis parameters and data.[14] Disclosure by OCME in this case included FST source code, project files, and database backups. I am unaware of any earlier direct review of FST independent of OCME. I have never seen formal, enumerated software requirement specifications for FST, against which FST could be tested for conformance.

33. The direct inspection of non-public FST materials allowed my company and Dr. Shapiro to identify components of FST which contained previously undisclosed behaviors, ultimately affecting reported likelihood ratio results under certain conditions. That FST contained undisclosed behaviors, coupled with a finding that the original FST program had been functionally modified *after* OCME had conducted its validation of FST led me to question whether FST should be considered validated at all. The original protective order preventing disclosure of our findings was vacated, which allowed for research into these behaviors to be conducted and conveyed to the greater scientific community through publication and conference presentations.[15,16]

---

[14] Adele A Mitchell et al. *Validation of a DNA Mixture Statistics Tool Incorporating Allelic Drop-out and Drop-in*, 6 Forensic Sci. Int'l: Genetics 749 (2012).

[15] J. Matthews, M. Babaeianjelodar, S. Lorenz, A. Matthews, M. Njie, N. Adams, D. Krane, J. Goldthwaite & C. Hughes, *The Right to Confront Your Accusers: Opening the Black Box of Forensic DNA Software*, Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), at 321–27 (2019).

[16] N. Adams, S. Lorenz, M. Babaeianjelodar, J. Matthews & D. Krane, *Quantifying the Impact of Post-validation Modifications to Forensic Statistical Tool*," 71st Annual Meeting of the Am. Acad.

34. Confidence in statistical outputs by probabilistic genotyping systems such as TrueAllele® should be informed by V&V outcomes. It is my opinion that V&V of a probabilistic genotyping system should be conducted in accordance with software industry standards before that system is used in casework.[17] While V&V processes described in ISO/IEC/IEEE Std 12207 and IEEE Std 1012 are intended to pervade the life of a software project, IEEE Std 1012-2012 recognizes, "Without a proactive approach, the anomalies and associated system changes are typically delayed to later in the program schedule, resulting in greater program costs and schedule delays." This does not suggest that V&V could be completed post-deployment (e.g. after casework use) but observes that even waiting until later stages of development *before* deployment can increase costs and time requirements.

35. Actual V&V efforts for low-integrity systems can be informal, bordering on non-existent. Therefore, it is difficult to estimate a lower bound for time and financial costs of V&V. High-integrity systems, such as the flight control software developed by IBM for the space shuttle, involved a verification team half the size of the development team and consuming 35% of the total development budget.[18] Stated expectations of IEEE Std 1012-2012 indicate that delayed V&V efforts would have increased these budgetary and timeline needs. It is reasonable to expect post-hoc V&V efforts to require even more resources. Delaying V&V efforts to the stage of criminal pre-trial proceedings is counter to IEEE Std 1012-2012 principles of ongoing V&V and would additionally subject V&V efforts to the time constraints of the court system.

---

of Forensic Sci. (Feb. 21, 2019) (presentation).

[17] N. Adams, R. Koppl, D. Krane, W. Thompson & S. Zabell, "Letter to the Editor-Appropriate Standards for Verification and Validation of Probabilistic Genotyping Systems," 63 J. Forensic Sci. 339–340 (2018).

[18] James E Tomayko, Nat'l Aeronautics & Space Admin., *Computers in Spaceflight: The NASA Experience* (1988), https://history.nasa.gov/computers/contents.html.

36. The April 2020 declaration by Dr. Perlin states at paragraph 46:

> TrueAllele has about 170,000 lines of computer source code, written by multiple programmers over two decades. The computer code is dense mathematical text. It can take hours for a person to read through even a few dozen lines of MATLAB to decipher what it does. Reading at ten lines per hour would entail eight and a half person-years to review all the source code.

37. I agree that a human reading 170,000 lines of code line-by-line would take an exceptionally long amount of time. At 50 lines per printed page, the source code would be 3,400 pages long. Thoroughly testing software can take even longer than reading source code since testing efforts require understanding the system's intended and actual behaviors.

38. While V&V of complex software systems involves reviews of source code and evaluating the extent of testing efforts, it does not involve reading the whole of a program's source code straight-through or testing every individual function.

39. The claim that TrueAllele® consists of 170,000 lines should call into question the government's statement about the disclosure of methods: "although the source code for TrueAllele is a secret, the methodology it employs and implements has been disclosed." The translation of any length of mathematical methods into 170,000 lines of code is non-trivial and subject to errors introduced through requirements definitions, design, and coding (construction/implementation). Necessary compromises between mathematical functions, however elegant, and the practical computational limitations of modern computers also create opportunities for edge cases that can cause aberrant program behaviors where none are intuitively expected and consequently are not appropriately guarded against.

40. The first step to a review of software verification and validation efforts is a simple inventory of what materials exist as well as how accessible and extensive they appear to be. General questions to answer at this stage include, "How well-defined is TrueAllele®? How were

certain tests selected to evaluate its behavior? How were those tests determined to be sufficient? How are changes to the program maintained over 20 years? Was it re-tested after modifications were made?"

41. Based on that initial inventory and domain-specific interests of biologists and statisticians involved in the inspection, particular inspections of materials can follow the inventory as a targeted audit. The software development materials and documentation requested in the April 29, 2020 amended subpoena are relevant to these types of inspections.

42. I estimate that an initial inventory of the TrueAllele® V&V-supportive materials that exist would take days to a few weeks, depending on how well-catalogued and centralized the materials are.

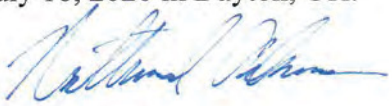
43. A more in-depth review could take longer and could reveal additional information about expected rates of unidentified defects, identification of extant defects, specific system behaviors including undisclosed behaviors, extents of testing efforts, and many other topics. If thorough documentation from earlier V&V efforts exists, an audit of those efforts could take weeks. If standard V&V tasks were not conducted, if documentation was not preserved, or if documentation is outdated or not well-centralized, I can offer no practical upper bound to a time estimate for a review. At a minimum, I would need 6-8 weeks to review and characterize a "worst-case scenario" for a program of the expected size and complexity of TrueAllele®.

V. Conclusion

44. Public materials describing the TrueAllele® software development process are not sufficient for determining that TrueAllele® has been developed in accordance with common software engineering practices or standards or that it could be or has been verified and validated against software engineering standards. Provision of materials supporting the TrueAllele®

software development and V&V processes would allow me to evaluate whether public claims of the reliable operation of TrueAllele® are supported by non-public materials.

July 16, 2020 in Dayton, OH.



Nathaniel Adams